



OIPE

*H2*

**RAW SEQUENCE LISTING**  
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002  
TIME: 11:36:15

Input Set : N:\Crf3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

**SEQUENCE LISTING**

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Schulein, Martin  
6 Andersen, Lene N.  
7 Lassen, Soren F.  
8 Kauppinen, Markus S.  
9 Lange, Lene  
10 Nielsen, Ruby I.  
11 Ihara, Michiko  
12 Takagi, Shinobu

14 (ii) TITLE OF INVENTION: Novel Endoglucanases

16 (iii) NUMBER OF SEQUENCES: 109

## 18 (iv) CORRESPONDENCE ADDRESS:

19 (A) ADDRESSEE: Novo Nordisk of North America, Inc.  
20 (B) STREET: 405 Lexington Avenue, 64th Floor  
21 (C) CITY: New York  
22 (D) STATE: New York  
23 (E) COUNTRY: United States of America  
24 (F) ZIP: 10174-6401

## 26 (v) COMPUTER READABLE FORM:

27 (A) MEDIUM TYPE: Floppy disk  
28 (B) COMPUTER: IBM PC compatible  
29 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
30 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

## 32 (vi) CURRENT APPLICATION DATA:

33 (A) APPLICATION NUMBER: US/10/007,521  
34 (B) FILING DATE: 10-Dec-2001

## 35 (C) CLASSIFICATION:

## 38 (vii) PRIOR APPLICATION DATA:

39 (A) APPLICATION NUMBER: US/08/651,136  
40 (B) FILING DATE: 21-MAY-1996

## 43 (viii) ATTORNEY/AGENT INFORMATION:

44 (A) NAME: Lambiris, Elias J.  
45 (B) REGISTRATION NUMBER: 33,728  
46 (C) REFERENCE/DOCKET NUMBER: 4366.200-US

## 48 (ix) TELECOMMUNICATION INFORMATION:

49 (A) TELEPHONE: 212-867-0123  
50 (B) TELEFAX: 212-878-9655

## 53 (2) INFORMATION FOR SEQ ID NO: 1:

55 (i) SEQUENCE CHARACTERISTICS:  
56 (A) LENGTH: 960 base pairs  
57 (B) TYPE: nucleic acid  
58 (C) STRANDEDNESS: single

**ENTERED**

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002

TIME: 11:36:15

Input Set : N:\Crf3\RULE60\10007521.raw

Output Set: N:\CRF3\02042002\J007521.raw

59 (D) TOPOLOGY: linear  
 61 (ii) MOLECULE TYPE: cDNA  
 63 (ix) FEATURE:  
 64 (A) NAME/KEY: CDS  
 65 (B) LOCATION: 113..787  
 67 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 69 AAAGAAAGGC TCTCTGCTGT CGTCGCTCTC GTCGCTCTCG TCGGCATCCT CCATCCGTCC 60  
 71 GCCTTGATA ACCCGCTCCC CGACTCAGTC AAGACGACGC ATACTTGGCA CC ATG 115  
 72 Met  
 73 1  
 75 CAT CTC TCC GCC ACC ACC GGG TTC CTC GCC CTC CCG GTC CTG GCC CTG 163  
 76 His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala Leu  
 77 5 10 15  
 79 GAC CAG CTC TCG GGC ATC GGC CAG ACG ACC CGG TAC TGG GAC TGC TGC 211  
 80 Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys Cys  
 81 20 25 30  
 83 AAG CCG AGC TGC GCC TGG CCC GGC AAG GGC CCC TCG TCT CCG GTG CAG 259  
 84 Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val Gln  
 85 35 40 45  
 87 GCC TGC GAC AAG AAC GAC AAC CCG CTC AAC GAC GGC GGC TCC ACC CGG 307  
 88 Ala Cys Asp Lys Asn Asn Pro Leu Asn Asp Gly Gly Ser Thr Arg  
 89 50 55 60 65  
 91 TCC GGC TGC GAC GCG GGC AGC GCC TAC ATG TGC TCC TCC CAG AGC 355  
 92 Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln Ser  
 93 70 75 80  
 95 CCC TGG GCC GTC AGC GAC GAG CTG TCG TAC GGC TGG GCG GCC GTC AAG 403  
 96 Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val Lys  
 97 85 90 95  
 99 CTC GCC GGC AGC TCC GAG TCG CAG TGG TGC TGC GCC TGC TAC GAG CTG 451  
 100 Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu Leu  
 101 100 105 110  
 103 ACC TTC ACC AGC GGG CCG GTC GCG GGC AAG AAG ATG ATT GTG CAG GCG 499  
 104 Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln Ala  
 105 115 120 125  
 107 ACC AAC ACC GGT GGC GAC CTG GGC GAC AAC CAC TTT GAC CTG GCC ATC 547  
 108 Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala Ile  
 109 130 135 140 145  
 111 CCC GGT GGC GGT GTC ATT TTC AAC GCC TGC ACC GAC CAG TAC GGC 595  
 112 Pro Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr Gly  
 113 150 155 160  
 115 GCT CCC CCG AAC GGC TGG GGC GAC CGC TAC GGC GGC ATC CAT TCC AAG 643  
 116 Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser Lys  
 117 165 170 175  
 119 GAA GAG TGC GAA TCC TTC CCG GAG GCC CTC AAG CCC GGC TGC AAC TGG 691  
 120 Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn Trp  
 121 180 185 190  
 123 CGC TTC GAC TGG TTC CAA AAC GCC GAC AAC CCG TCG GTC ACC TTC CAG 739  
 124 Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe Gln  
 125 195 200 205

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002  
TIME: 11:36:15

Input Set : N:\Crf3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

127 GAG GTG GCC TGC CCG TCG GAG CTC ACG TCC AAG AGC GGC TGC TCC CGT	787
128 Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser Arg	
129 210 215 220 225	
131 TAAGAGGGAA GAGAGGGGGC TTGAAGGACC GAAAGATTCA ACCTCTGCTC CTGCTGGGGA	847
133 AGCTCGGGCG CGAGTGTGAA ACTGGTGTAA ATATTGTGGC ACACACAAGC TACTACAGTC	907
135 CGTCTCGCCG TCCGGCTAAC TAGCCTTGCT GCGGATCTGT CCAAAAAAAA AAA	960
<b>138 (2) INFORMATION FOR SEQ ID NO: 2:</b>	
140 (i) SEQUENCE CHARACTERISTICS:	
141 (A) LENGTH: 225 amino acids	
142 (B) TYPE: amino acid	
143 (D) TOPOLOGY: linear	
145 (ii) MOLECULE TYPE: protein	
147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:	
149 Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala	
150 1 5 10 15	
152 Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys	
153 20 25 30	
155 Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val	
156 35 40 45	
158 Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr	
159 50 55 60	
161 Arg Ser Gly Cys Asp Ala Gly Ser Ala Tyr Met Cys Ser Ser Gln	
162 65 70 75 80	
164 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val	
165 85 90 95	
167 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu	
168 100 105 110	
170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln	
171 115 120 125	
173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala	
174 130 135 140	
176 Ile Pro Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr	
177 145 150 155 160	
179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser	
180 165 170 175	
182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn	
183 180 185 190	
185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe	
186 195 200 205	
188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser	
189 210 215 220	
191 Arg	
192 225	
195 (2) INFORMATION FOR SEQ ID NO: 3:	
197 (i) SEQUENCE CHARACTERISTICS:	
198 (A) LENGTH: 894 base pairs	
199 (B) TYPE: nucleic acid	
200 (C) STRANDEDNESS: single	
201 (D) TOPOLOGY: linear	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002  
TIME: 11:36:15

Input Set : N:\CrF3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

203	(ii) MOLECULE TYPE: cDNA	
205	(ix) FEATURE:	
206	(A) NAME/KEY: CDS	
207	(B) LOCATION: 1..891	
209	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:	
211	ATG CAT CTC TCC GCC ACC ACC GGG TTC CTC GCC CTC CCG GTC CTG GCC	48
212	Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala	
W--> 213	230 235 240	
215	CTG GAC CAG CTC TCG GGC ATC GGC CAG ACG ACC CGG TAC TGG GAC TGC	96
216	Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys	
W--> 217	245 250 255	
219	TGC AAG CCG AGC TGC GCC TGG CCC GGC AAG GGC CCC TCG TCT CCG GTG	144
220	Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val	
W--> 221	260 265 270	
223	CAG GCC TGC GAC AAG AAC GAC AAC CCG CTC AAC GAC GGC GGC TCC ACC	192
224	Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr	
W--> 225	275 280 285	
227	CGG TCC GGC TGC GAC GCG GGC AGC GCC TAC ATG TGC TCC TCC CAG	240
228	Arg Ser Gly Cys Asp Ala Gly Ser Ala Tyr Met Cys Ser Ser Gln	
W--> 229	290 295 300 305	
231	AGC CCC TGG GCC GTC AGC GAC GAG CTG TCG TAC GGC TGG GCG GCC GTC	288
232	Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val	
W--> 233	310 315 320	
235	AAG CTC GCC GGC AGC TCC GAG TCG CAG TGG TGC TGC GCC TGC TAC GAG	336
236	Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu	
W--> 237	325 330 335	
239	CTG ACC TTC ACC AGC GGG CCG GTC GCG GGC AAG AAG ATG ATT GTG CAG	384
240	Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln	
W--> 241	340 345 350	
243	GCG ACC AAC ACC GGT GGC GAC CTG GGC GAC AAC CAC TTT GAC CTG GCC	432
244	Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala	
W--> 245	355 360 365	
247	ATC CCC GGT GGC GGT GTC GGT ATT TTC AAC GCC TGC ACC GAC CAG TAC	480
248	Ile Pro Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr	
W--> 249	370 375 380 385	
251	GGC GCT CCC CCG AAC GGC TGG GGC GAC CGC TAC GGC GGC ATC CAT TCC	528
252	Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser	
W--> 253	390 395 400	
255	AAG GAA GAG TGC GAA TCC TTC CCG GAG GCC CTC AAG CCC GGC TGC AAC	576
256	Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn	
W--> 257	405 410 415	
259	TGG CGC TTC GAC TGG TTC CAA AAC GAC AAC CCG TCG GTC ACC TTC	624
260	Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe	
W--> 261	420 425 430	
263	CAG GAG GTG GCC TGC CCG TCG GAG CTC ACG TCC AAG AGC GGC TGC TCC	672
264	Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser	
W--> 265	435 440 445	
267	CGT CCC TCC AGC AGC ACC AGC TCT CCG GTC AAC CAG CCT ACC AGC ACC	720
268	Arg Pro Ser Ser Thr Ser Ser Pro Val Asn Gln Pro Thr Ser Thr	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002  
TIME: 11:36:15

Input Set : N:\Crf3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

W--> 269	450	455	460	465	
271	AGC ACC ACG TCC ACC ACC TCG AGC CCG CCA GTC CAG CCT ACG				768
272	Ser Thr Thr Ser Thr Ser Thr Ser Pro Pro Val Gln Pro Thr				
W--> 273	470	475	480		
275	ACT CCC AGC GGC TGC ACT GCT GAG AGG TGG GCT CAG TGC GGC GGC AAT				816
276	Thr Pro Ser Gly Cys Thr Ala Glu Arg Trp Ala Gln Cys Gly Gly Asn				
W--> 277	485	490	495		
279	GGC TGG AGC GGC TGC ACC ACC TGC GTC GCT GGC AGC ACT TGC ACG AAG				864
280	Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys				
W--> 281	500	505	510		
283	ATT AAT GAC TGG TAC CAT CAG TGC CTG TAG				894
284	Ile Asn Asp Trp Tyr His Gln Cys Leu				
W--> 285	515	520			
288	(2) INFORMATION FOR SEQ ID NO: 4:				
290	(i) SEQUENCE CHARACTERISTICS:				
291	(A) LENGTH: 297 amino acids				
292	(B) TYPE: amino acid				
293	(D) TOPOLOGY: linear				
295	(ii) MOLECULE TYPE: protein				
297	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:				
299	Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala				
300	1 5 10 15				
302	Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys				
303	20 25 30				
305	Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val				
306	35 40 45				
308	Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr				
309	50 55 60				
311	Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln				
312	65 70 75 80				
314	Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val				
315	85 90 95				
317	Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu				
318	100 105 110				
320	Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Met Ile Val Gln				
321	115 120 125				
323	Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala				
324	130 135 140				
326	Ile Pro Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr				
327	145 150 155 160				
329	Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser				
330	165 170 175				
332	Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn				
333	180 185 190				
335	Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe				
336	195 200 205				
338	Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser				
339	210 215 220				
341	Arg Pro Ser Ser Ser Thr Ser Pro Val Asn Gln Pro Thr Ser Thr				

VERIFICATION SUMMARY DATE: 02/04/2002  
PATENT APPLICATION: US/10/007,521 TIME: 11:36:16

Input Set : N:\Crf3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

L:33 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:34 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:213 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:217 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:221 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:225 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:229 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:233 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:237 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:241 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:245 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:249 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:253 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:257 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:261 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:265 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:269 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:273 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:277 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:281 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:285 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:375 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:379 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:383 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:387 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:391 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:395 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:399 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:403 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:407 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:411 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:415 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:419 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:423 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:427 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:431 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:435 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:439 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:443 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:447 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:451 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5  
L:544 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:548 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:552 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:556 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:560 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:564 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:568 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7

VERIFICATION SUMMARY  
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002  
TIME: 11:36:16

Input Set : N:\Crf3\RULE60\10007521.raw  
Output Set: N:\CRF3\02042002\J007521.raw

L:572 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:576 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:580 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:584 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:3513 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:79  
L:3529 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:80  
L:3545 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:81  
L:3561 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:82  
L:3577 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:83  
L:3907 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:105  
L:3922 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:106  
L:3937 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:107  
L:3952 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:108  
L:3967 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:109